

# semana 10

Abigail Ramos

13/10/2022

```
## Loading required package: splines
## Loading required package: RcmdrMisc
## Loading required package: car
## Loading required package: carData
## Loading required package: sandwich
## Loading required package: effects
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

## La interfaz R-Commander sólo funciona en sesiones interactivas
##
## Attaching package: 'Rcmdr'

## The following object is masked from 'package:base':
##
##      errorCondition
```

## Intervalo de confianza para la media en una población normal con varianza conocida

### Supuesto Práctico 1

El archivo empleados.xls nos informa de la edad, altura, peso, sexo y posesión de coche de 100 empleados de una empresa. Suponiendo la normalidad de la variable Altura, calcular el intervalo de confianza sobre la altura media poblacional a un 95% de confianza, sabiendo que la varianza poblacional es 6.

Importar datos de Excel a R-Commander

```
> empleados <-
+   readXL("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/empleados.xls"
+   rownames=FALSE, header=TRUE, na="", sheet="Respuestas", stringsAsFactors=TRUE)
```

A continuación guardamos el archivo (Datos/Conjunto de datos activo/Guardar el conjunto de datos activo) como fichero Empleados.RData.

```
> save("empleados",
+   file="C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/empleados.R.Data")
```

Introducimos en R los datos, que proporciona el enunciado, relativos al nivel de significación y la varianza poblacional de la variable.

```
> alpha<-0.05
> varianza<-7.5
```

Calculamos por separado cada uno de los elementos restantes que necesitamos para obtener el intervalo de confianza.

```
> n <- nrow(empleados)
> media <- mean(empleados$Altura)
>
> cuantil<- qnorm(1-alpha/2)
```

Por último, calculamos los extremos inferior y superior del intervalo de acuerdo

```
> lim_inferior<-media -cuantil * sqrt(varianza) / sqrt(n)
> lim_inferior
```

```
[1] 176.4605
```

```
> lim_superior<- media + cuantil * sqrt(varianza) / sqrt(n)
> lim_inferior
```

```
[1] 176.4605
```

Por lo que el intervalo de confianza que buscamos es (176.4605, 177.5395).

## Intervalo de confianza para la media en una población normal con varianza desconocida

### Supuesto Práctico 2

Considerando el conjunto de datos de empleados.xls y asumiendo que la variable que mide la altura de los empleados sigue una distribución Normal con varianza desconocida. Calcular un intervalo de confianza a un nivel de confianza del 90% para la altura media poblacional.

Accedemos al menú Test t para una muestra, seleccionando en el menú principal: Estadísticos/ Medias/ Test t para una muestra.

```
> with(empleados, (t.test(Altura, alternative='two.sided', mu=0.0, conf.level=.90)))
```

```
One Sample t-test
```

```
data:  Altura
t = 214.41, df = 98, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 175.6292 178.3708
sample estimates:
mean of x
 177
```

En la parte superior izquierda se muestra una lista con todas las variables cuantitativas del archivo de datos que son susceptibles de ser contrastadas, de la cual debemos elegir exclusivamente una.

El intervalo de confianza pedido es (175.6292, 178.3708).

## Intervalo de confianza para la proporción

### Supuesto Práctico 3

A partir del conjunto de datos de empleados.xls, obtener un intervalo de confianza al 95% para la proporción de empleados varones en la población.

Solución

Accedemos al menú Test de proporciones para una muestra de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para una muestra.

En la pestaña Datos del cuadro de diálogo, se muestra una lista con todas las variables cualitativas que pueden utilizarse en este tipo de contrastes, de entre las cuales tenemos que elegir una. En este caso elegimos Sexo.

```
> local({
+   .Table <- xtabs(~ Sexo , data= empleados )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95, correct=FALSE)
+ })
```

Frequency counts (test is for first level):

Sexo

Hombre Mujer

87 12

1-sample proportions test without continuity correction

data: rbind(.Table), null probability 0.5

X-squared = 56.818, df = 1, p-value = 4.78e-14

alternative hypothesis: true p is not equal to 0.5

95 percent confidence interval:

0.7999934 0.9292845

sample estimates:

p

0.8787879

En este ejemplo, las dos posibles opciones para la variable Sexo son “Hombre” y “Mujer”, por lo que la primera de las categorías para R-Commander es “Hombre”. Dado que la hipótesis que se ha planteado se ha hecho sobre los hombres no es necesario hacer ninguna modificación.

Si, por el contrario, la hipótesis del problema se hubiera planteado sobre las mujeres, deberíamos hacer una recodificación previa de la variable para situar la categoría “Mujer” como la primera.

En la segunda parte se muestran los resultados del contraste de hipótesis, que analizaremos en la práctica sobre Contrastes de Hipótesis.

Por lo que el intervalo de confianza, a un nivel de confianza del 95% para la proporción de empleados varones en la población es (0.7999934, 0.9292845).

## Intervalo de confianza para el cociente de varianzas en dos poblaciones normales independientes

### Supuesto Práctico 4

Continuando con los datos del archivo empleados.xls y asumiendo que el peso en hombres y el peso en mujeres se distribuyen según distribuciones normales con medias y varianzas desconocidas. Calcular un intervalo de confianza a un nivel de confianza del 95% para el cociente de varianzas en ambas poblaciones. ¿Puede asumirse que ambas varianzas son iguales?

Accedemos a Test F para dos varianzas, de R-Commander, seleccionando en el menú principal: Estadísticos/Varianzas/ Test F para dos varianzas

La pestaña Datos muestra dos listas de variables.

- La lista de Grupos, muestra todas las variables cualitativas del fichero de datos. En esta lista tenemos que seleccionar cuál es la variable que nos va a dividir la muestra de observaciones en dos submuestras independientes. En nuestro caso, al tratar con hombres y mujeres, seleccionamos el Sexo del empleado como variable de agrupación.
- La lista de Variable explicativa, muestra todas las variables cuantitativas del fichero de datos. Señalamos la variable principal sobre la cual se va a llevar a cabo el contraste (en nuestro caso, Peso).

```
> Tapply(Peso ~ Sexo, var, na.action=na.omit, data=empleados) # variances by group
```

```
      Hombre      Mujer
166.99626  92.06061
```

```
> var.test(Peso ~ Sexo, alternative='two.sided', conf.level=.95, data=empleados)
```

```
      F test to compare two variances
```

```
data:  Peso by Sexo
F = 1.814, num df = 86, denom df = 11, p-value = 0.2752
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6112226 3.8937784
sample estimates:
ratio of variances
      1.813982
```

Esta salida, muestra que el intervalo de confianza para el cociente de las varianzas es (0.6112226, 3.8937784).

La interpretación del intervalo de confianza puede servirnos para concluir acerca de la igualdad de las varianzas. En este ejemplo, dicho intervalo es (0.6112226, 3.8937784) que, como podemos comprobar incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 95% se puede suponer que el cociente entre las dos varianzas puede tomar el valor 1 o, lo que es lo mismo, que las dos varianzas son iguales.

Una vez se ha determinado la igualdad (o desigualdad) de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias propiamente dicho.

**a) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas pero supuestas iguales**

**b) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas, distintas y tamaños muestrales grandes**

Supuesto Práctico 5

Sabiendo que las varianzas son iguales (Supuesto práctico 4), obtener un intervalo de confianza al 95% para la diferencia del peso medio entre hombres y mujeres. ¿Puede suponerse que el peso medio entre hombres y mujeres es igual?

Para obtener el Test t para muestras independientes con R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para muestras independientes.

La pestaña Datos muestra dos listas de variables. Como ya se ha comentado con anterioridad, en la lista de la izquierda (Grupos) tenemos que escoger la variable a partir de la cual se formarán los dos grupos de observaciones, seleccionamos Sexo. En la de la derecha (Variable explicada) seleccionamos la variable cuya diferencia de medias en las poblaciones queremos estudiar, seleccionamos Peso.

```
> t.test(Peso~Sexo, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=empleados)
```

Welch Two Sample t-test

data: Peso by Sexo

t = 3.8302, df = 17.057, p-value = 0.001333

alternative hypothesis: true difference in means between group Hombre and group Mujer is not equal to 0

95 percent confidence interval:

5.329686 18.394452

sample estimates:

mean in group Hombre mean in group Mujer

76.19540

64.33333

Se puede afirmar que el intervalo de confianza a un 95% de confianza para la diferencia de las medias del peso medio para hombres y mujeres es (4.167581, 19.556557). Como el 0 no está dentro de este intervalo, no tenemos suficiente evidencia muestral para decir que el peso medio de hombres y mujeres sea el mismo.

## Intervalo de confianza para la diferencia de medias en dos poblaciones normales relacionadas

### Supuesto Práctico 6

Se desea evaluar la eficacia de un fármaco para la reducción del nivel de glucosa en pacientes. Para ello, se selecciona una muestra de 10 pacientes a los que se les mide su nivel de glucosa en sangre antes y después del suministro del medicamento. Los resultados aparecen recogidos en la siguiente tabla:

### Solución

En primer lugar, debemos crear un nuevo conjunto de datos con la información que nos proporciona la tabla. El conjunto de datos estará formado por dos variables con los niveles de glucosa antes y después de la aplicación del fármaco. Organizamos los datos en un fichero .txt (supuesto6.txt)

A continuación seleccionamos en el menú principal: Datos/Importar datos/desde archivo de texto

```
> supuesto6 <-
+ read.table("C:/Users/abbyc/Desktop/Ciclo II 2022/Análisis estadístico con R/curso-R-2022/supuesto6.txt",
+ header=TRUE, stringsAsFactors=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
```

A continuación, accedemos al menú Test t para datos relacionados en R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para datos relacionados

Es importante destacar que, a diferencia del caso de muestras independientes, cuando trabajamos con muestras pareadas no necesitamos una variable de agrupación, sino que debemos seleccionar las dos variables a analizar de forma separada.

En la pestaña Opciones personalizamos el contraste conforme al problema que estemos resolviendo. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado.

```
> with(supuesto6, (t.test(Antes, Después, alternative='two.sided', conf.level=.90,
+ paired=TRUE)))
```

Paired t-test

data: Antes and Después

t = -2.1581, df = 5, p-value = 0.08338

alternative hypothesis: true mean difference is not equal to 0

90 percent confidence interval:

```

-2.9972387 -0.1027613
sample estimates:
mean difference
-1.55

```

En la segunda parte de los resultados se incluye el intervalo de confianza al 90% para la diferencia de las medias de la variable glucosa, que es (-9.290501, 16.790501). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambos niveles medios de glucosa son iguales.

## Intervalo de confianza para la diferencia de dos proporciones

### Supuesto Práctico 7

A partir del conjunto de datos empleados.xls, obtener un intervalo de confianza al 85% para la diferencia entre la proporción de empleados hombres y mujeres que tienen coche. ¿Pueden considerarse ambas proporciones iguales?

### Solución

Dado que la hipótesis se ha planteado sobre la proporción de hombres y mujeres que tienen coche, es necesario recodificar la variable coche. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

En esta pantalla:

- Se selecciona Coche en Variables a recodificar
- En Nuevo nombre o prefijo para variables múltiples recodificadas hemos puesto como nombre de la variable: coche\_rec
- En la ventana Introducir directrices de recodificación: “No”=”2No” y “Sí”=”1Sí”

```

> empleados <- within(empleados, {
+   coche_rec <- Recode(Coche, '"No"="2No"; "Sí"="1Sí";', as.factor=TRUE)
+ })

```

Accedemos al menú Test de proporciones para dos muestras de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para dos muestras

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Sexo, ya que distinguimos entre hombres y mujeres) y de la segunda, la variable de interés (que es, si el empleado tiene coche, coche\_rec).

En la pestaña Opciones indicamos el nivel de confianza propuesto en el enunciado, 85%

```

> library(abind, pos=17)

> local({ .Table <- xtabs(~Sexo+coche_rec, data=empleados)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.85, correct=FALSE)
+ })

```

Percentage table:

	coche_rec			
Sexo	1Sí	2No	Total	Count
Hombre	52.9	47.1	100	87
Mujer	50.0	50.0	100	12

## 2-sample test for equality of proportions without continuity correction

```
data: .Table
X-squared = 0.03492, df = 1, p-value = 0.8518
alternative hypothesis: two.sided
85 percent confidence interval:
 -0.1928653  0.2503366
sample estimates:
   prop 1   prop 2 
0.5287356 0.5000000
```

El intervalo de confianza pedido al 85% de confianza es (-0.2729805, 0.3304517). El 0 está dentro de este intervalo, por lo que podemos concluir que las proporciones de hombre y mujeres que tienen coche coinciden.